FIG. 1

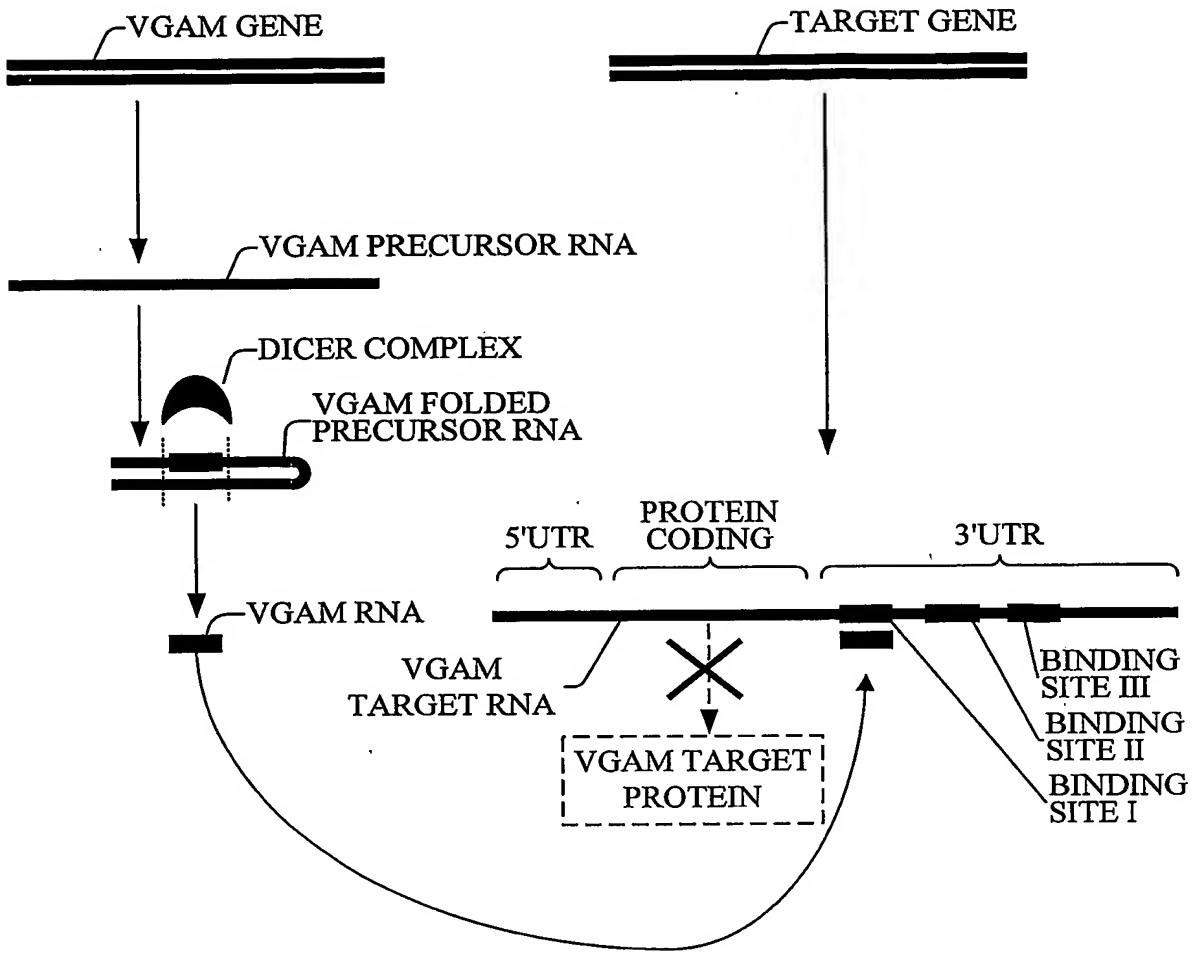

VGAM GENE
TARGET GENE
VGAM PRECURSOR RNA
DICER COMPLEX
VGAM FOLDED PRECURSOR RNA
5'UTR
PROTEIN CODING
3'UTR
VGAM RNA
VGAM TARGET RNA
VGAM TARGET PROTEIN
BINDING SITE III
BINDING SITE II
BINDING SITE I

FIG. 2

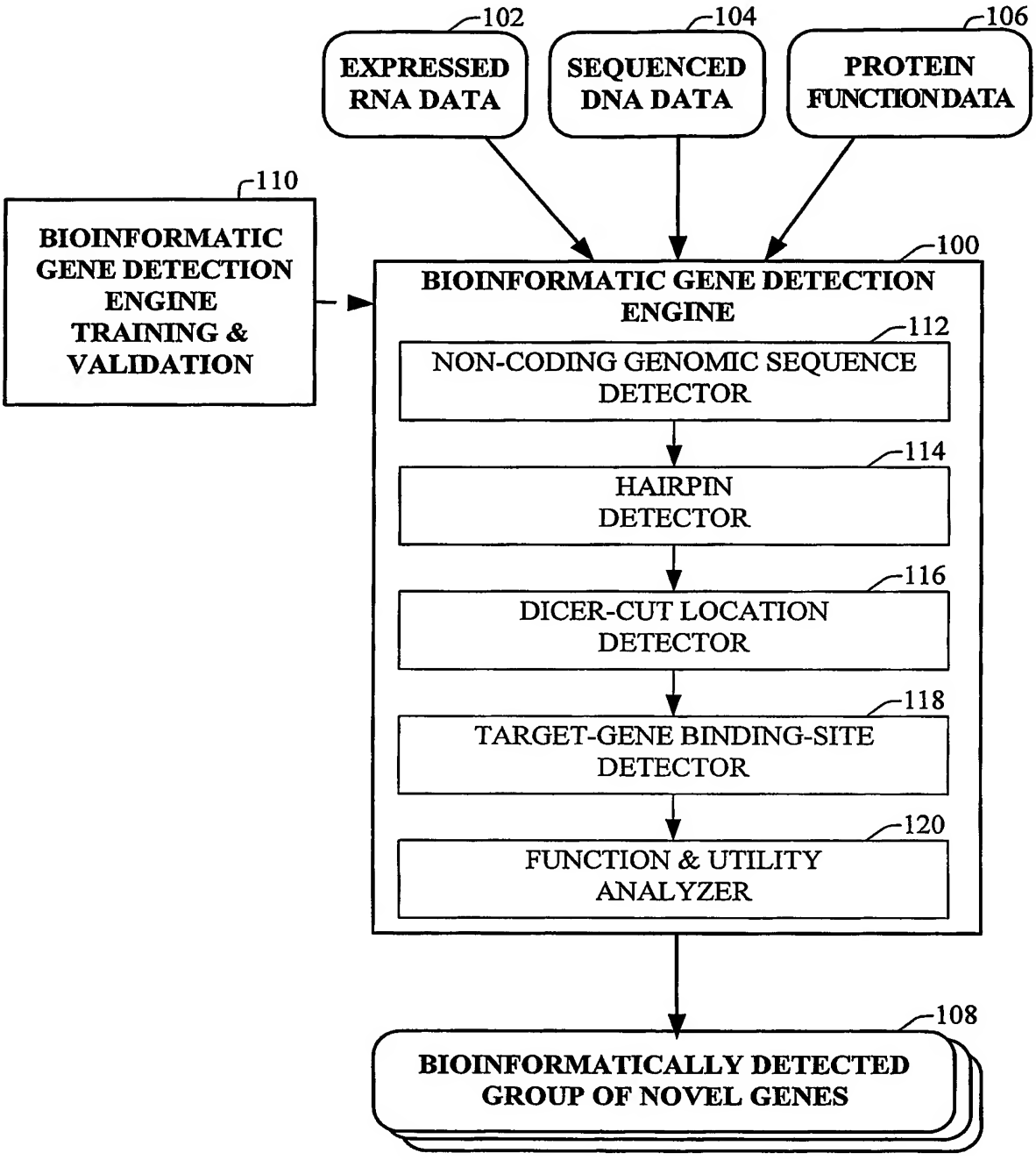

102
EXPRESSED RNA DATA
104
SEQUENCED DNA DATA
106
PROTEIN FUNCTIONDATA
110
BIOINFORMATIC GENE DETECTION ENGINE TRAINING & VALIDATION
100
BIOINFORMATIC GENE DETECTION ENGINE
112
NON-CODING GENOMIC SEQUENCE DETECTOR
114
HAIRPIN DETECTOR
116
DICER-CUT LOCATION DETECTOR
118
TARGET-GENE BINDING-SITE DETECTOR
120
FUNCTION & UTILITY ANALYZER
108
BIOINFORMATICALLY DETECTED GROUP OF NOVEL GENES

FIG. 3

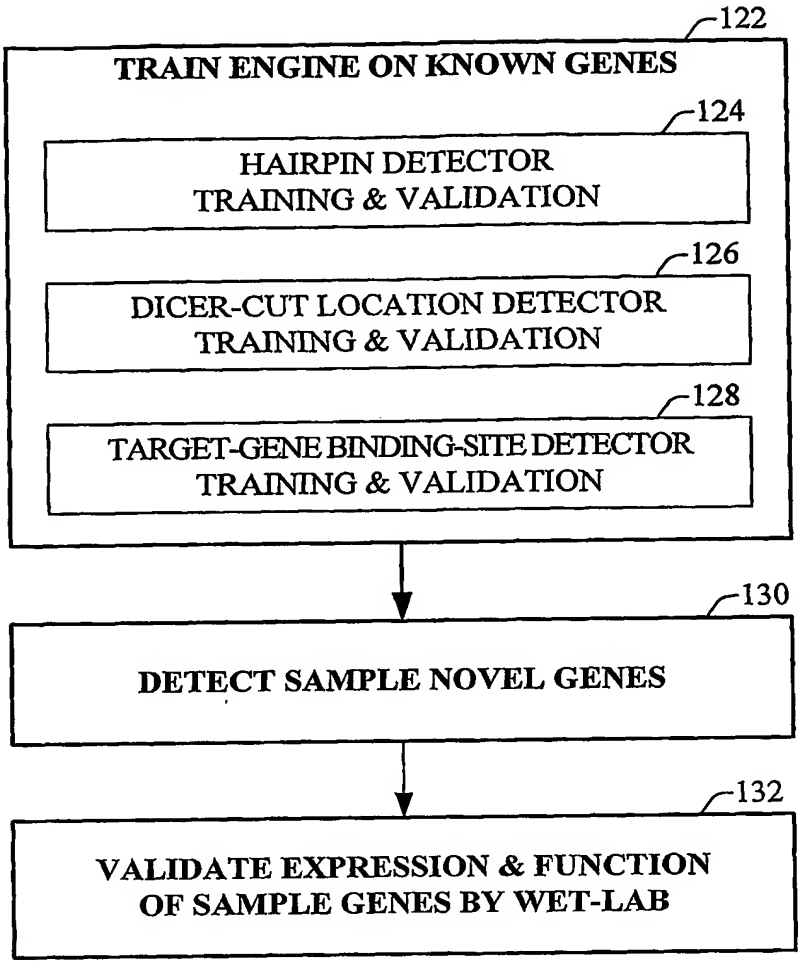

122
TRAIN ENGINE ON KNOWN GENES
124
HAIRPIN DETECTOR TRAINING & VALIDATION
126
DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION
128
TARGET-GENE BINDING-SITE DETECTOR TRAINING & VALIDATION
130
DETECT SAMPLE NOVEL GENES
132
VALIDATE EXPRESSION & FUNCTION OF SAMPLE GENES BY WET-LAB

FIG. 4A

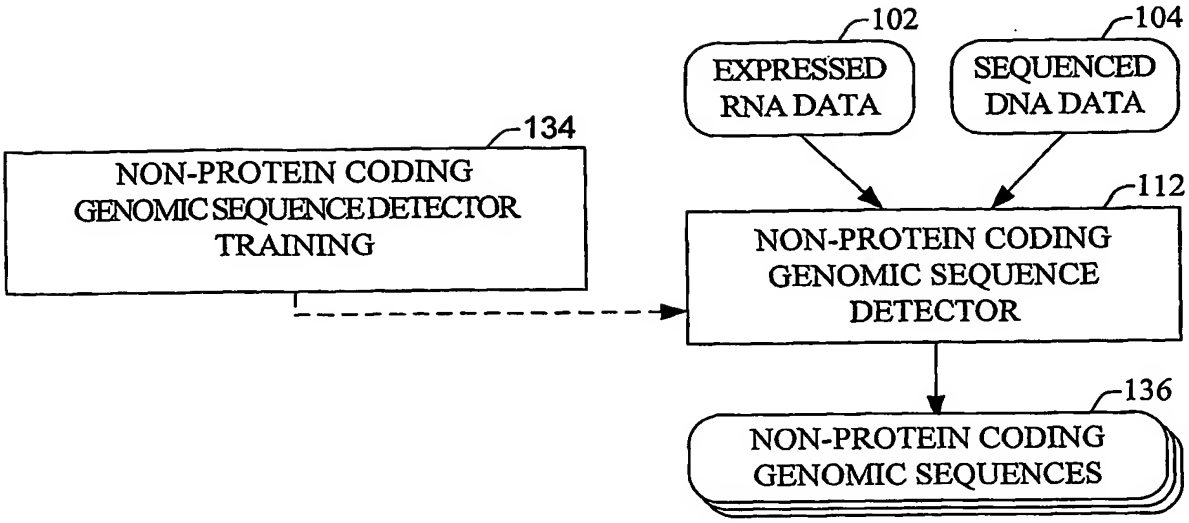

102
104
EXPRESSED RNA DATA
SEQUENCED DNA DATA
134
NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR TRAINING
112
NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR
136
NON-PROTEIN CODING GENOMIC SEQUENCES

FIG. 4B

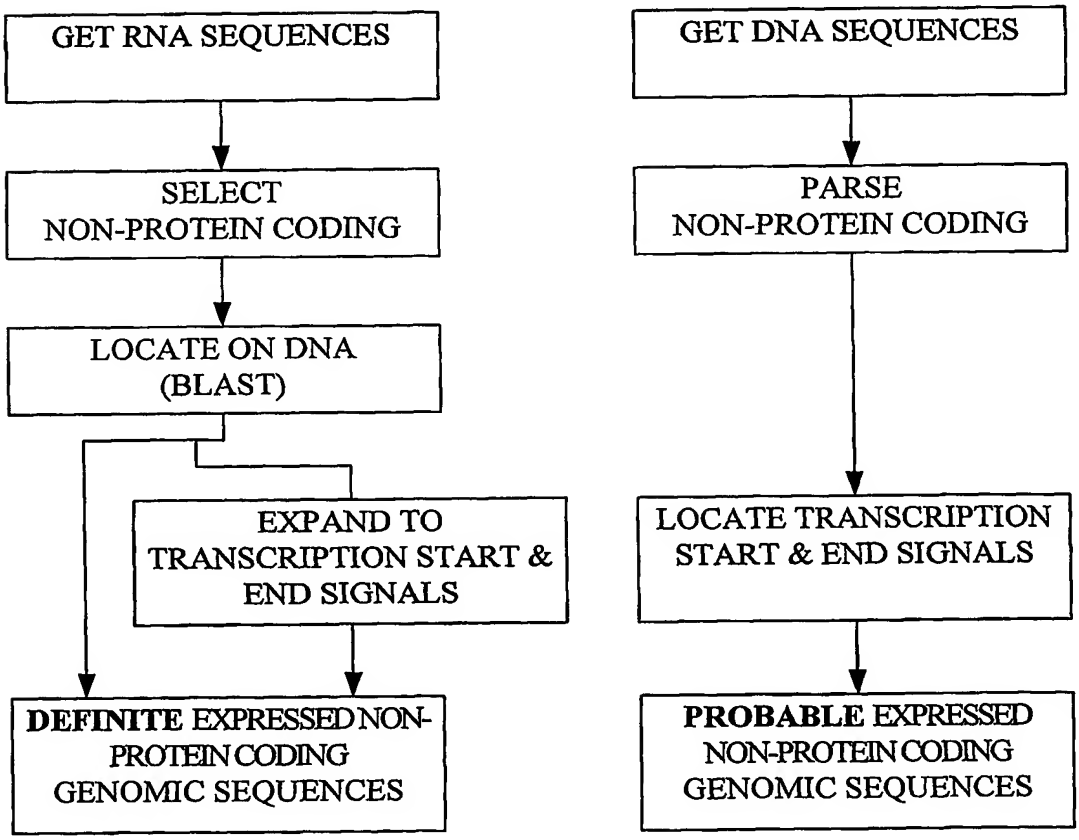

GET RNA SEQUENCES
SELECT NON-PROTEIN CODING
LOCATE ON DNA (BLAST)
EXPAND TO TRANSCRIPTION START & END SIGNALS
DEFINITE EXPRESSED NON-PROTEIN CODING GENOMIC SEQUENCES
GET DNA SEQUENCES
PARSE NON-PROTEIN CODING
LOCATE TRANSCRIPTION START & END SIGNALS
PROBABLE EXPRESSED NON-PROTEIN CODING GENOMIC SEQUENCES

FIG. 5A

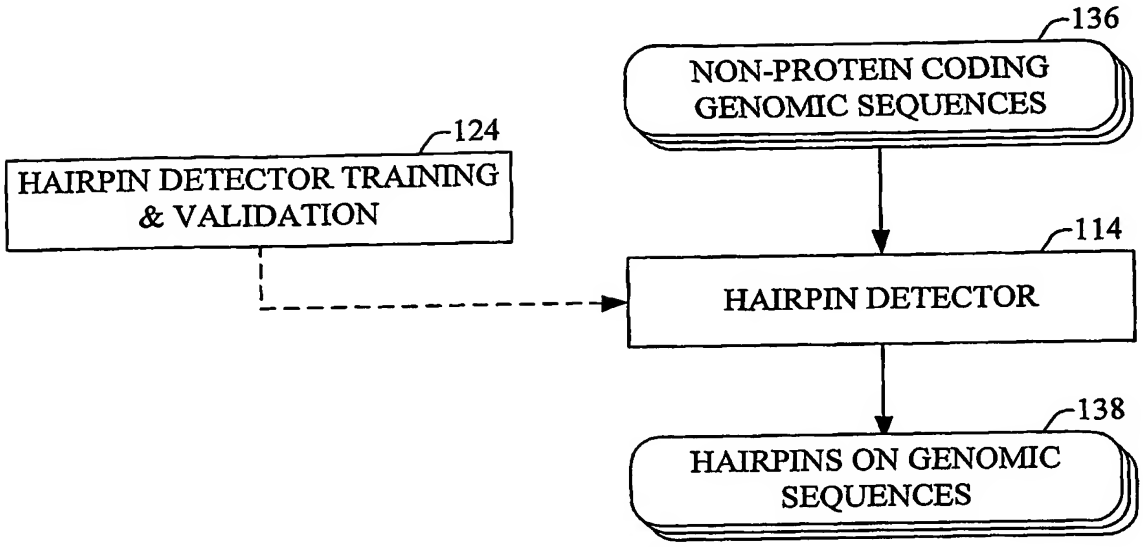

136
NON-PROTEIN CODING GENOMIC SEQUENCES
124
HAIRPIN DETECTOR TRAINING & VALIDATION
114
HAIRPIN DETECTOR
138
HAIRPINS ON GENOMIC SEQUENCES

FIG. 5B

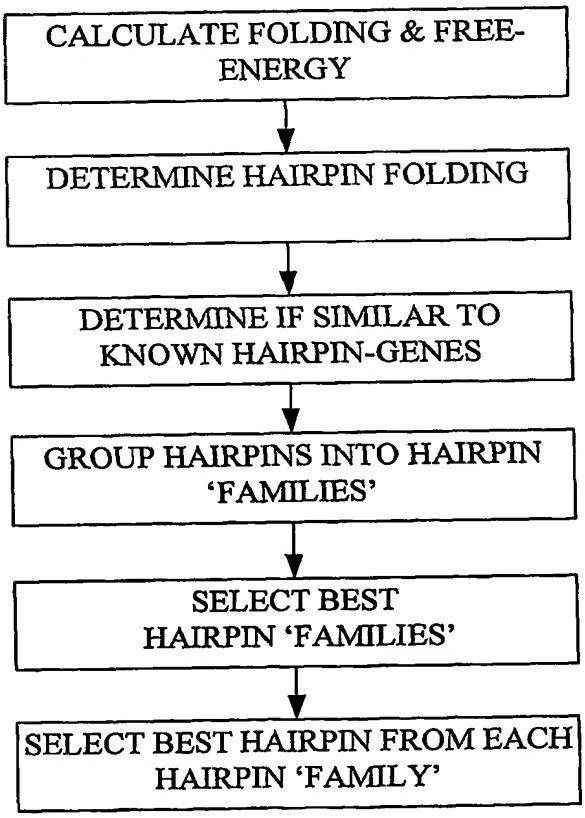

CALCULATE FOLDING & FREE-ENERGY
DETERMINE HAIRPIN FOLDING
DETERMINE IF SIMILAR TO KNOWN HAIRPIN-GENES
GROUP HAIRPINS INTO HAIRPIN 'FAMILIES'
SELECT BEST HAIRPIN 'FAMILIES'
SELECT BEST HAIRPIN FROM EACH HAIRPIN 'FAMILY'

FIG. 6A

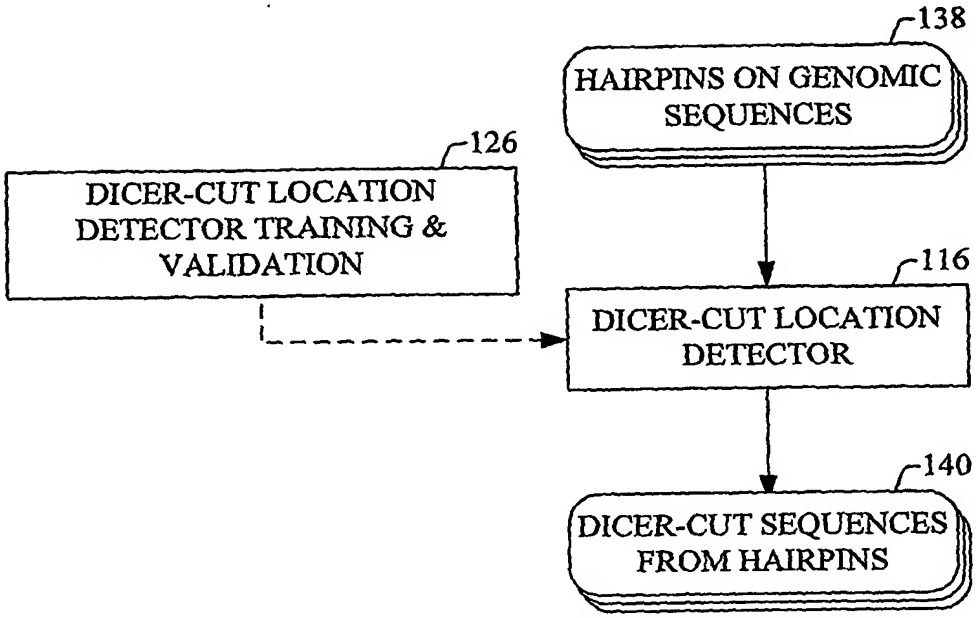

138
HAIRPINS ON GENOMIC SEQUENCES
126
DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION
116
DICER-CUT LOCATION DETECTOR
140
DICER-CUT SEQUENCES FROM HAIRPINS

FIG. 6B

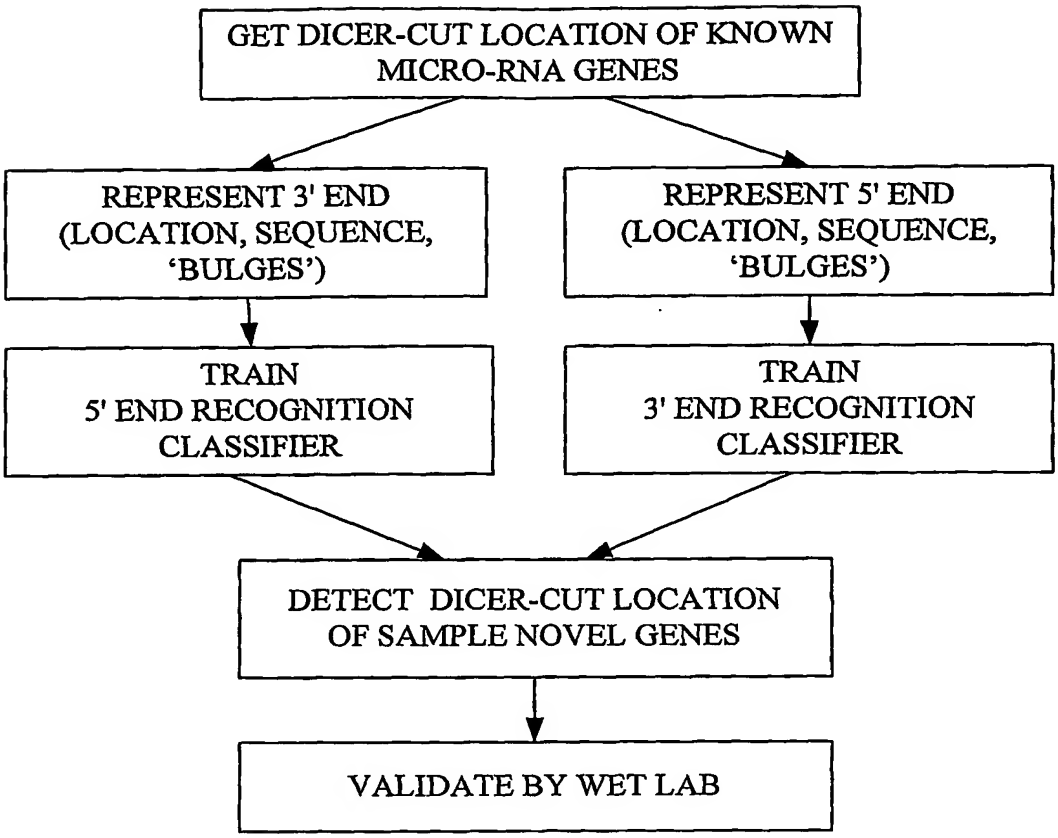

GET DICER-CUT LOCATION OF KNOWN MICRO-RNA GENES
REPRESENT 3' END (LOCATION, SEQUENCE, 'BULGES')
REPRESENT 5' END (LOCATION, SEQUENCE, 'BULGES')
TRAIN 5' END RECOGNITION CLASSIFIER
TRAIN 3' END RECOGNITION CLASSIFIER
DETECT DICER-CUT LOCATION OF SAMPLE NOVEL GENES
VALIDATE BY WET LAB

FIG. 6C

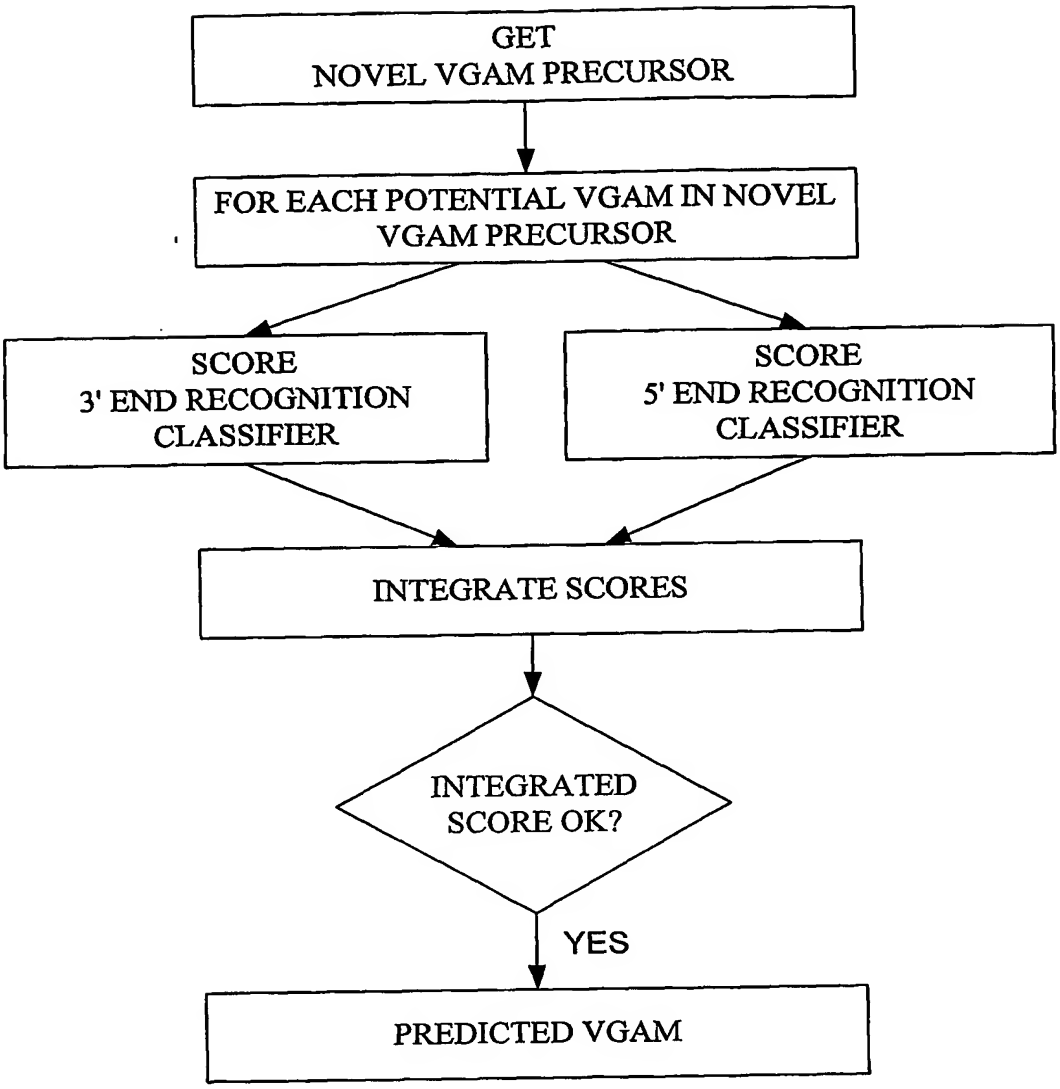

GET
NOVEL VGAM PRECURSOR
FOR EACH POTENTIAL VGAM IN NOVEL VGAM PRECURSOR
SCORE
3' END RECOGNITION
CLASSIFIER
SCORE
5' END RECOGNITION
CLASSIFIER
INTEGRATE SCORES
INTEGRATED
SCORE OK?
YES
PREDICTED VGAM

FIG. 7A

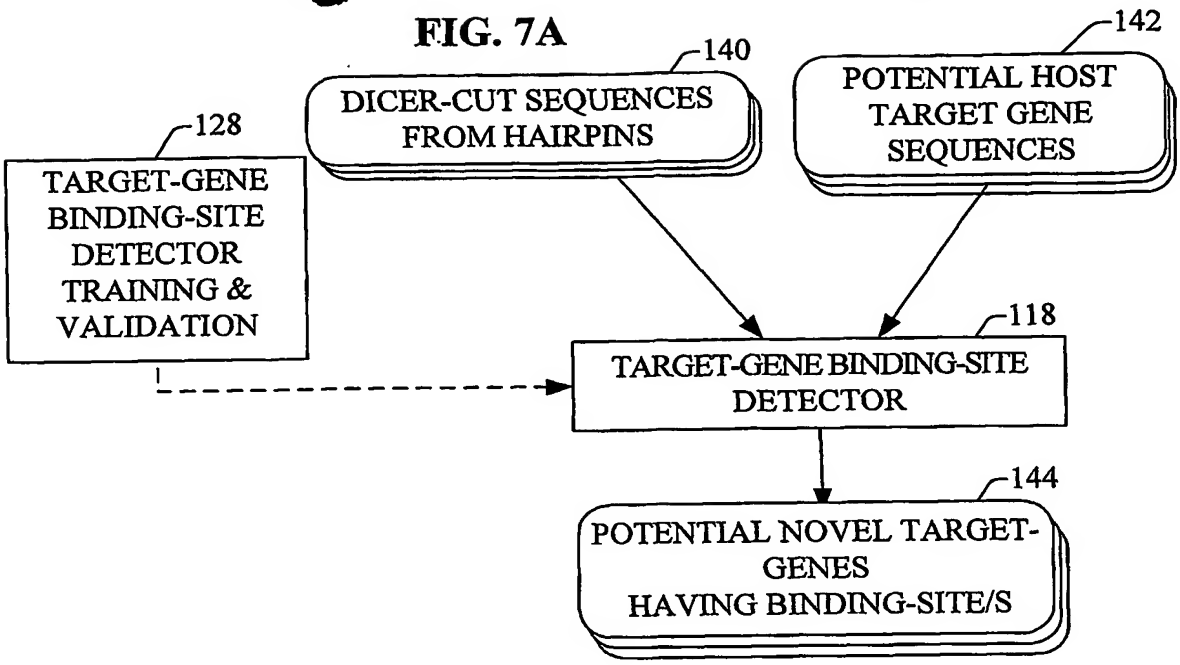

140
DICER-CUT SEQUENCES FROM HAIRPINS
142
POTENTIAL HOST TARGET GENE SEQUENCES
128
TARGET-GENE BINDING-SITE DETECTOR TRAINING & VALIDATION
118
TARGET-GENE BINDING-SITE DETECTOR
144
POTENTIAL NOVEL TARGET-GENES HAVING BINDING-SITE/S

FIG. 7B

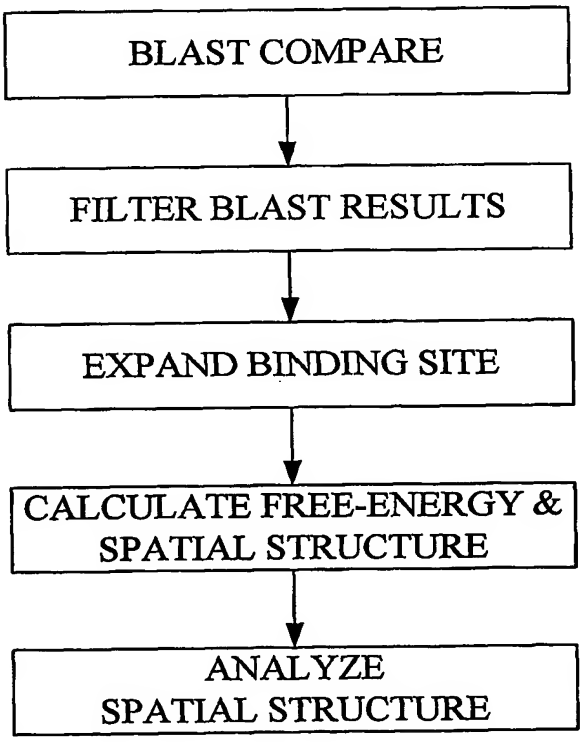

BLAST COMPARE
FILTER BLAST RESULTS
EXPAND BINDING SITE
CALCULATE FREE-ENERGY & SPATIAL STRUCTURE
ANALYZE SPATIAL STRUCTURE

FIG. 8

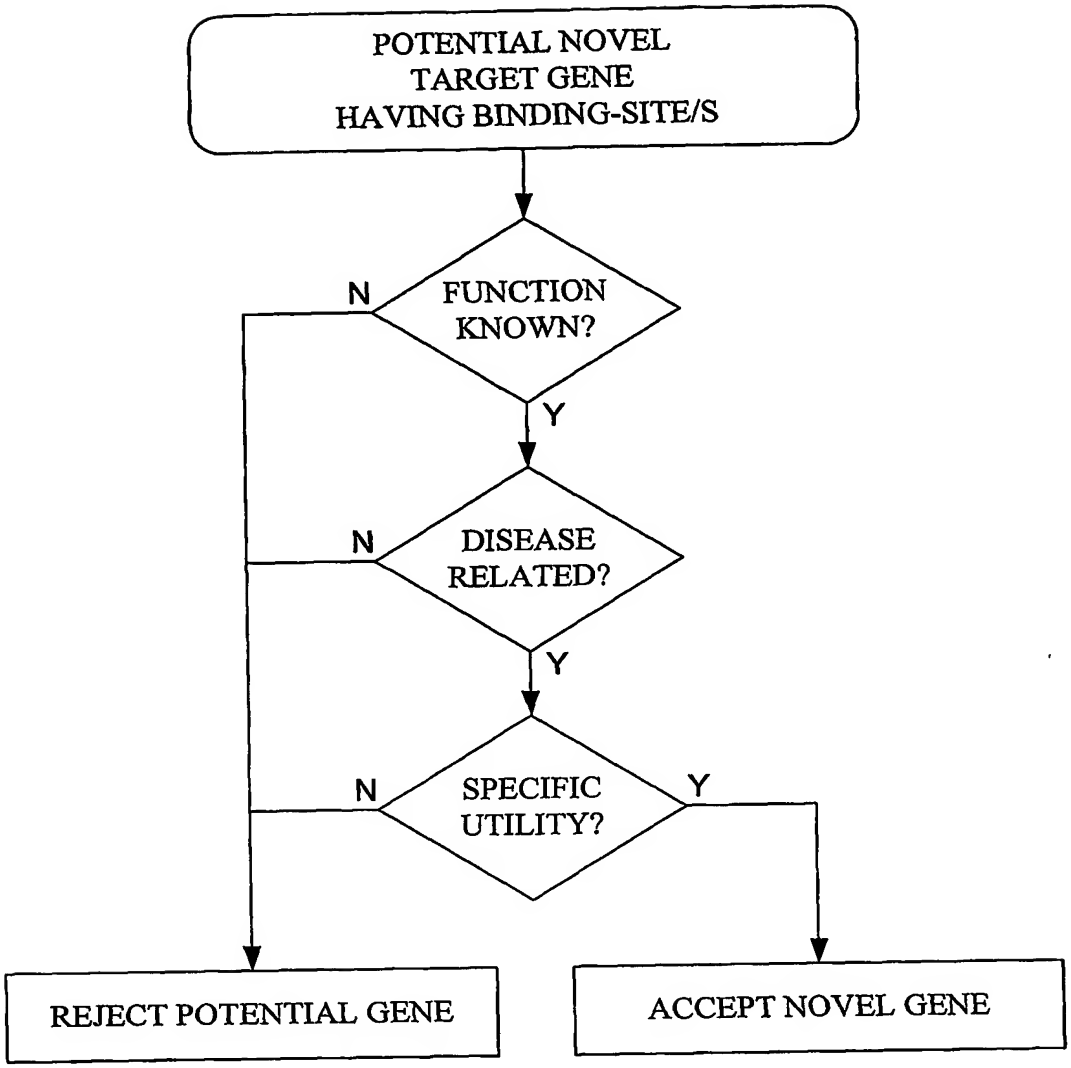

POTENTIAL NOVEL TARGET GENE HAVING BINDING-SITE/S
FUNCTION KNOWN?
N
Y
DISEASE RELATED?
N
Y
SPECIFIC UTILITY?
N
Y
REJECT POTENTIAL GENE
ACCEPT NOVEL GENE

FIG. 9

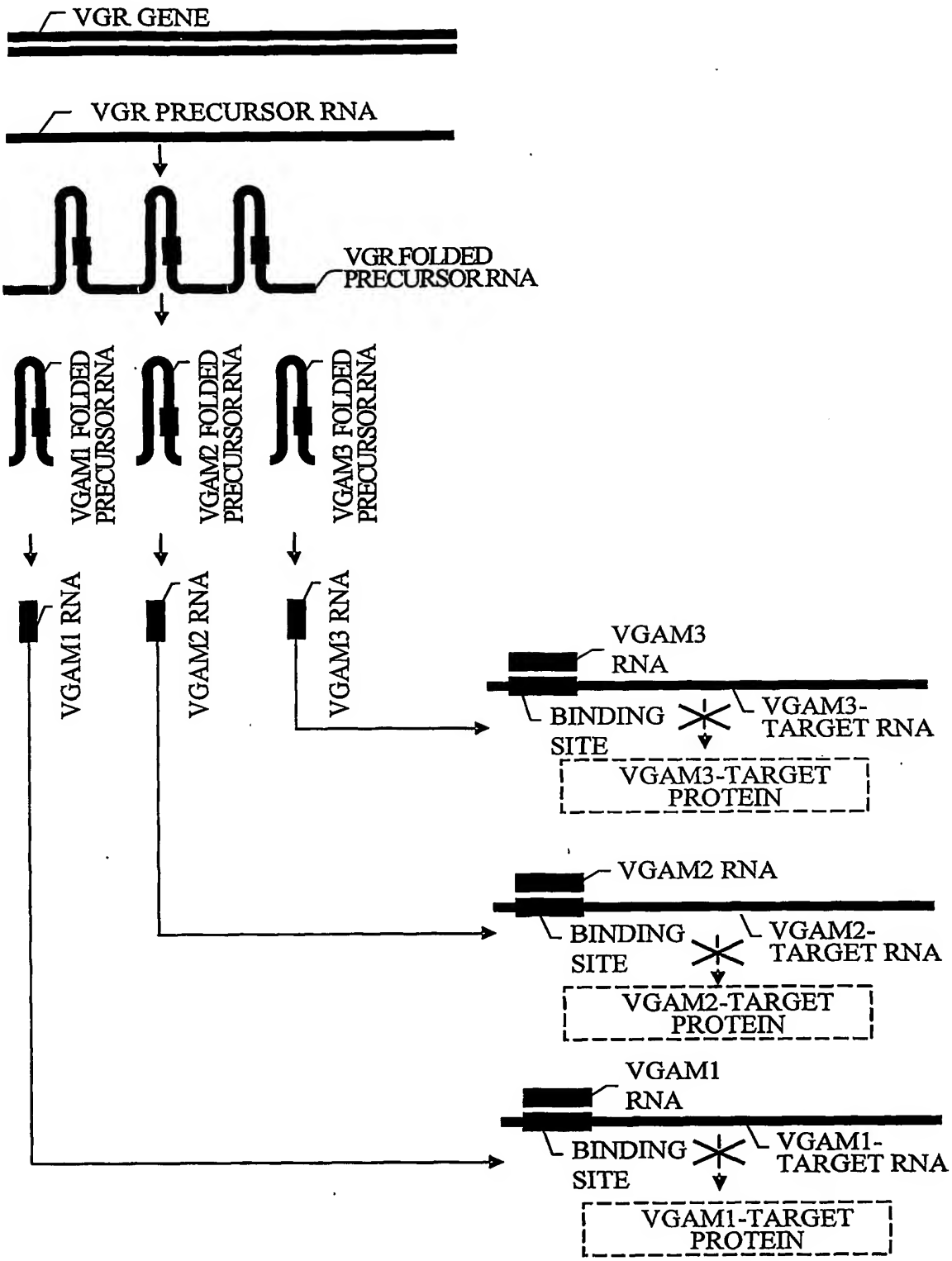

VGR GENE
VGR PRECURSOR RNA
VGR FOLDED PRECURSOR RNA
VGAM1 FOLDED PRECURSOR RNA
VGAM2 FOLDED PRECURSOR RNA
VGAM3 FOLDED PRECURSOR RNA
VGAM1 RNA
VGAM2 RNA
VGAM3 RNA
VGAM3 RNA
BINDING SITE
VGAM3-TARGET RNA
VGAM3-TARGET PROTEIN
VGAM2 RNA
BINDING SITE
VGAM2-TARGET RNA
VGAM2-TARGET PROTEIN
VGAM1 RNA
BINDING SITE
VGAM1-TARGET RNA
VGAM1-TARGET PROTEIN

FIG. 10

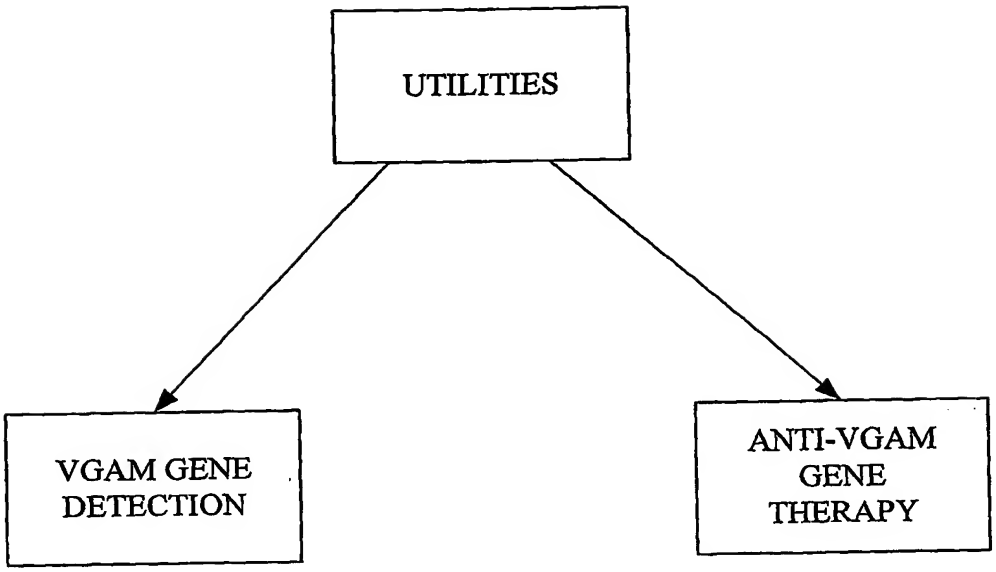


FIG. 11A

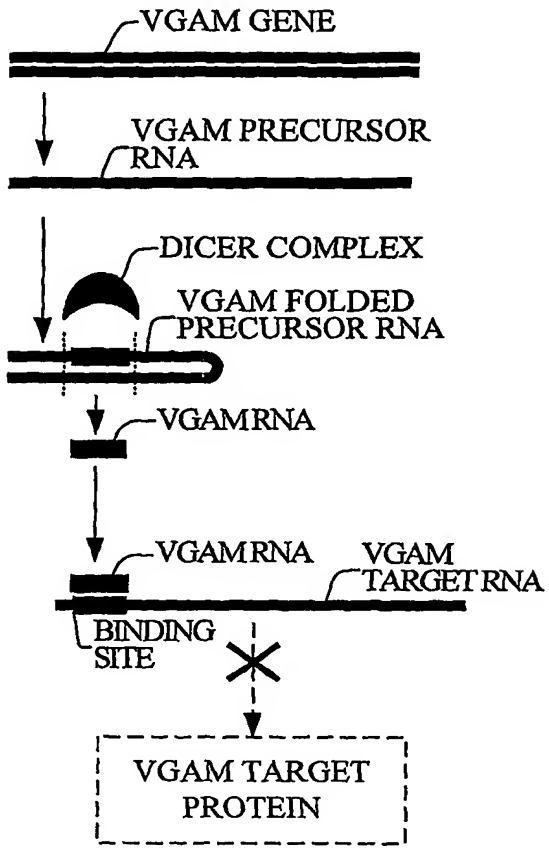


FIG. 11B

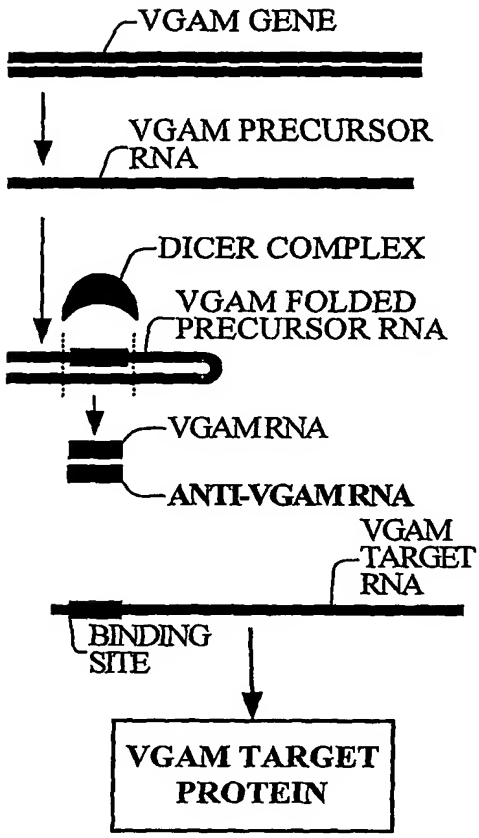
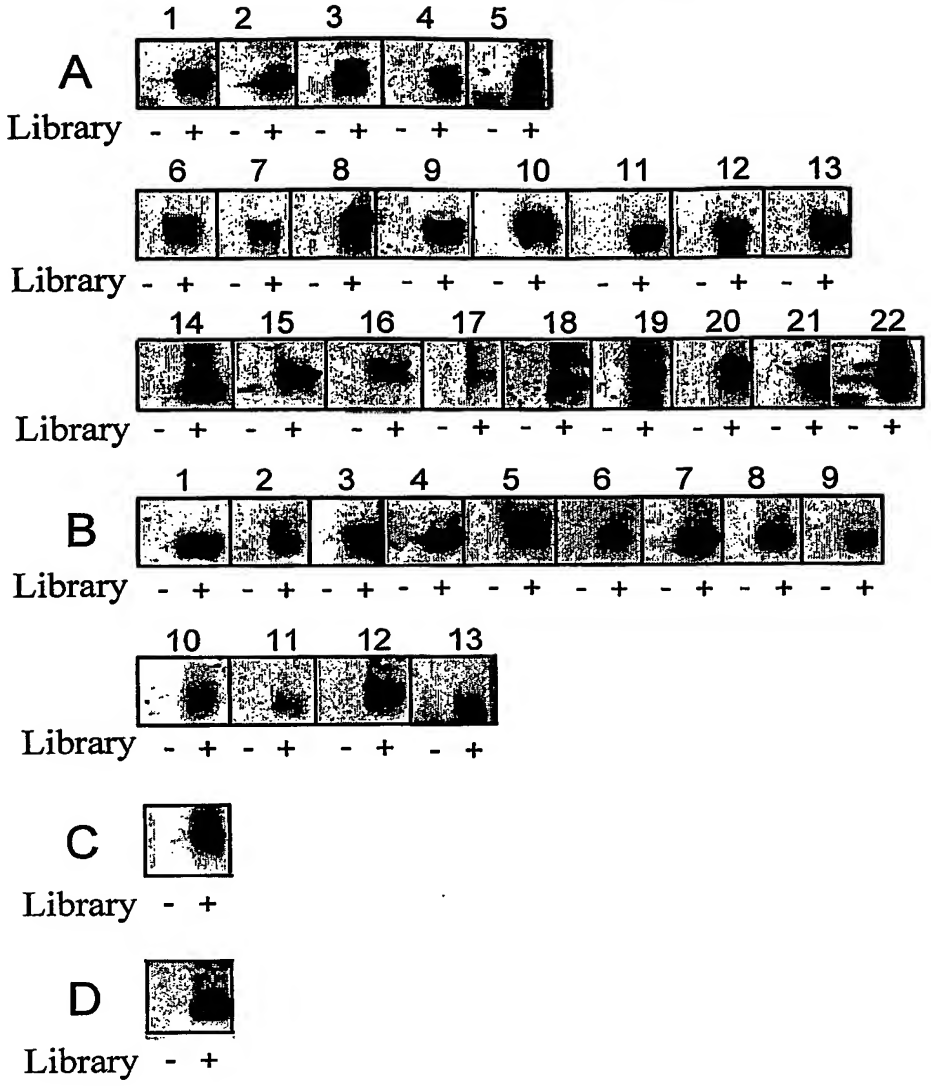


FIG. 12

Detection Accuracy Group	Number of Published MIRs	Precision on hairpin mixture	Lab Validation of Human GAMs			Novel VGAMs of Present Invention
			Sent	Positive	% success	
A	228	76%	101	37	37 %	20
B	135	41%	56	13	23 %	342
C	27	18%	7	1	14 %	212
D	20	10%	4	1	25 %	322
Overall	410	44%	168	52	31 %	896

FIG. 13


1 2 3 4 5
A
Library - + - + - + - + - +
6 7 8 9 10 11 12 13
Library - + - + - + - + - + - + - + - +
14 15 16 17 18 19 20 21 22
Library - + - + - + - + - + - + - + - + - +
1 2 3 4 5 6 7 8 9
B
Library - + - + - + - + - + - + - + - + - +
10 11 12 13
Library - + - + - + - +
C
Library - +
D
Library - +

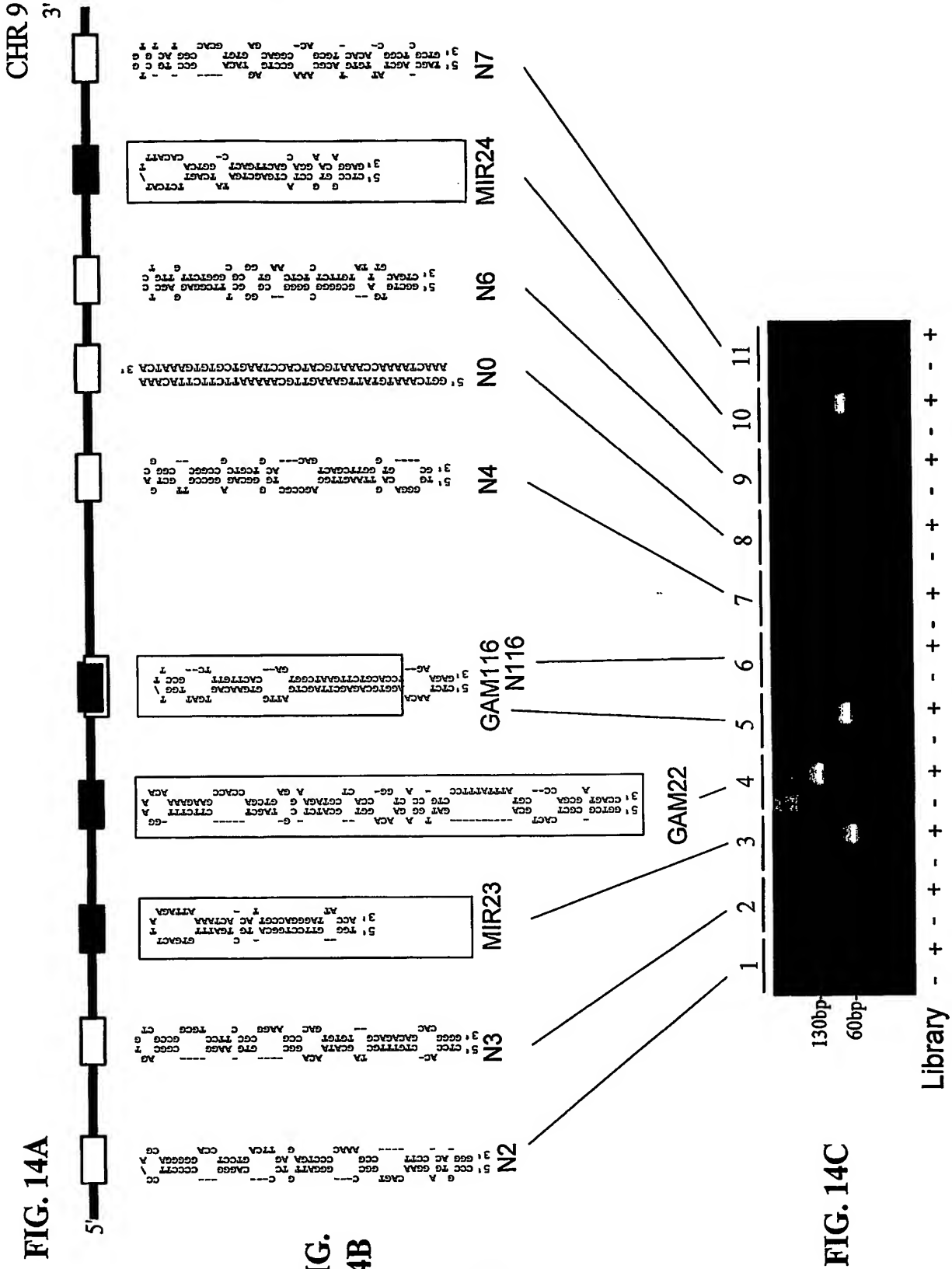


FIG. 14A

FIG. 14B

FIG. 14C

EST72223 (705 nt.)

FIG. 15A

EST72223 sequence:

CCCTTATTAGAGGATTCTGCTCATGCCAGGGTGAGGTAGTAAGTTGTATTG
TTGTGGGGTAGGGATATTAGGCCCCAATTAGAAGATAACTATACAACT MIR98
TACTACTTTCCCTGGTGTGTGGCATATTCACACTTAGTCTTAGCAGTGTTGCC
TCCATCAGACAAAGTTGTAGATGTTCCTTGGATAATTTGGACTGGAAGAAAAGA
GACATGGAAGGGGACAGATGGTGTTTAGGGTGAGGCAGATGTCATTATAAAGT
GACTTGTCTTTCATTAATTGGAGCATATAATTATTTTACCTTTGGGCATGAACTC
ATTTTGCTATTCTTCAACTGTGTAATGATTGCATTTTATTAGTAATAGAACAGGA
ATGTGTGCAAGGGAATGGAAAGCATACTTTAAGAATTTTGGGCCAGGCGCGGT
GGTTCATGCCTGTAATCCCAGCATTTTTGGGAGGCCGAGGCGGGTGGATCAC
CTGAGGTCAGGAGTTCGAGACCAACCTGGCCAACACGGCGAAACCCCGCCTC
TACTCAAATACAAAAATTAGCCAGGCTTGGTGACACTCGCCTGTGGTCCCAGC
TACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAAGTGGAG GAM25
GCTTCAGTGAGCTGAGAACACGCCACTGCACTCCAGTCCTGGGCAAC
AGAGCAAGACTCTGTCTCAGGAAAAAAAAAG

FIG. 15B

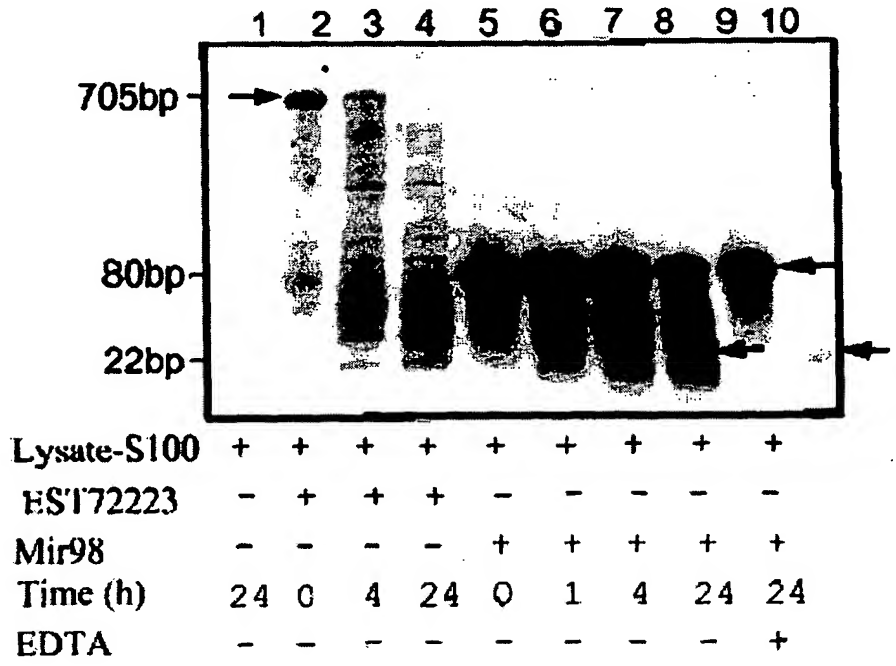


FIG. 15D

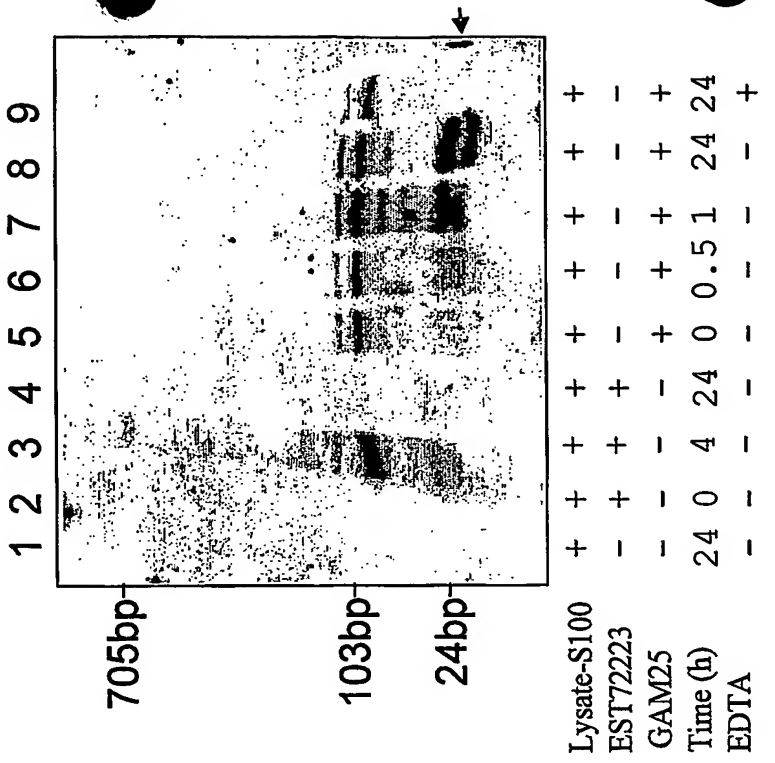


FIG. 15C

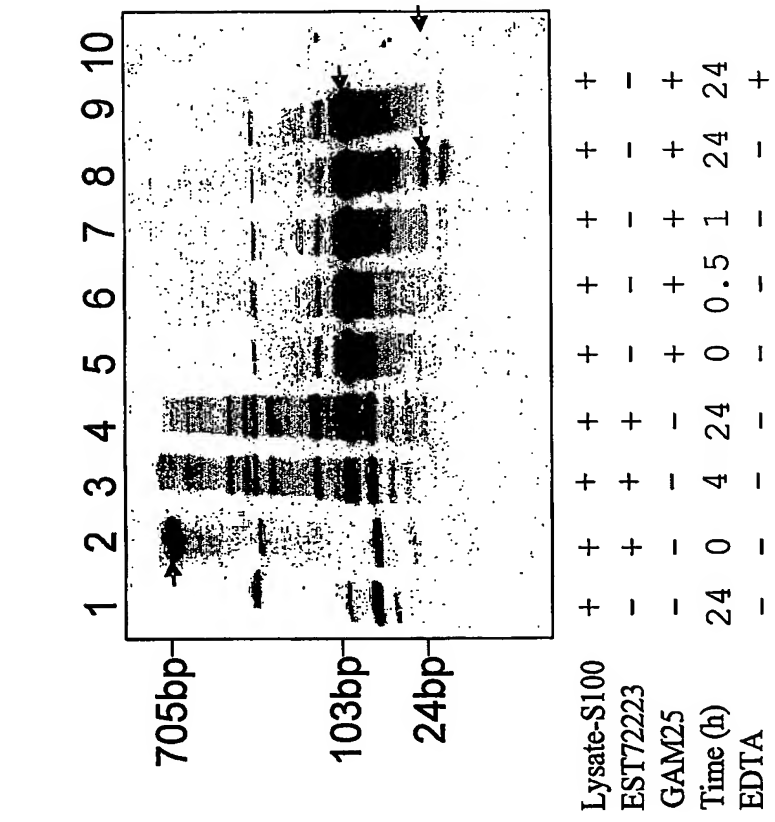


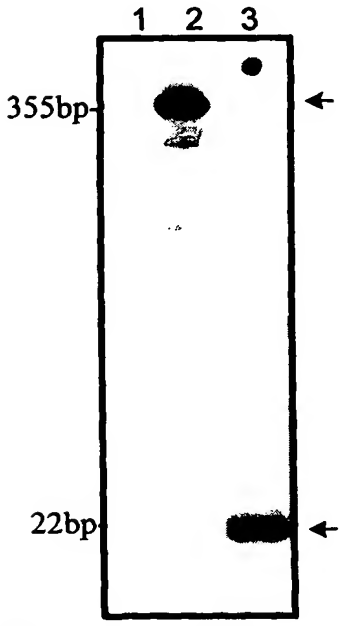
FIG. 16A

5'UTR SEQUENCE (5' TO 3') OF HIV-1(U5-R)

GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACT
AGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTA
GTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTT
TTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACCTGAAAG
CGAAAGGGAAACCAGAGGAGCTCTCTCGA**CGCAGGACTCGGCTTGCTGAA**
GCGCGCACGGCAAGAGGCGAGGGGCGGGCGACTGGTGAGTACGCCAAAAA
TTTTGACTAGCGGAGGCTAGAAGGAGAGAG

FIG. 16B

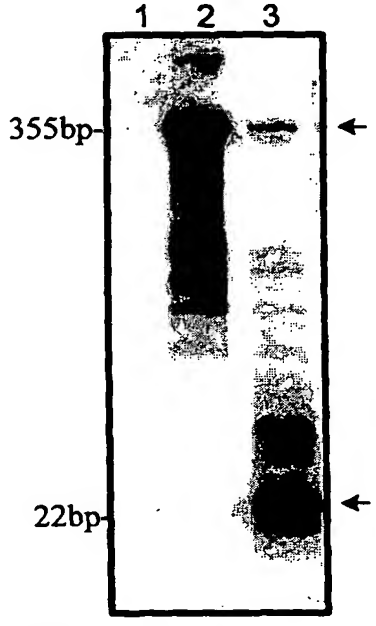
VGAM2032.2



Lysate s-100	+	+	+
U5R	-	+	+
Time(h)	24	0	24

FIG. 16C

VGAM3249.1



Lysate s-100	+	+	+
U5R	-	+	+
Time(h)	24	0	24

SUBSTITUTE SHEET (RULE 26)

FIG.16D

TRANSCRIPT SEQUENCE (5' TO 3')

ACAGGGCTGCTATTAACAAGAG
ATGGTGGTAATAGCAA**CAATGA**
GTCCGAGATCTTCAGACCTGG
AGGAGGAGATATGAGGGACAA
TTGGAGAAGTGAATTATATAAA
TATAAAGTAGTAAAAATTGAACC
ATTAGGAGTAGCACCCACCAAG
GCAAAGAGAA

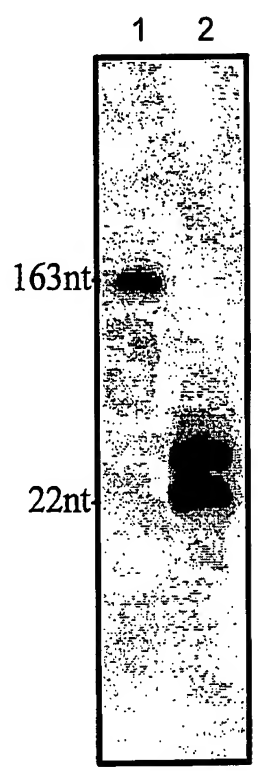


FIG.16E

TRANSCRIPT SEQUENCE (5' TO 3')

TGGGAAGCCCTCAAATATTGGTGG
AATCTCCTACAGTATTGGAGTCAG
GAACTAAAGAATAG**TGCTGTTAG**
CTTGCTCAATGCCACAGCCATAG
CAGTAGCTGAGGGGACAGATAG
GGTTATAGAAGTAGTACAAGGAG
CTTGTAGAGCTATTCGCCACATACC
TAGAAGAATAAGACAGGGCTTGGA
AAGGATTTTGCT

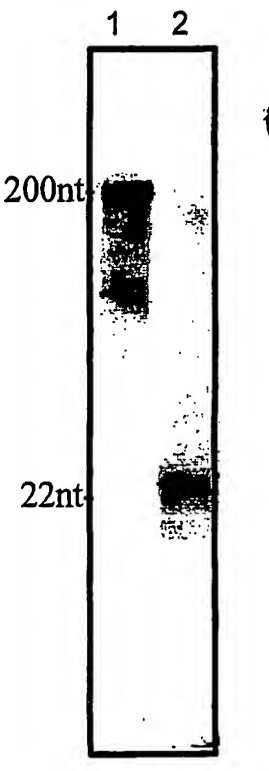


FIG.17A

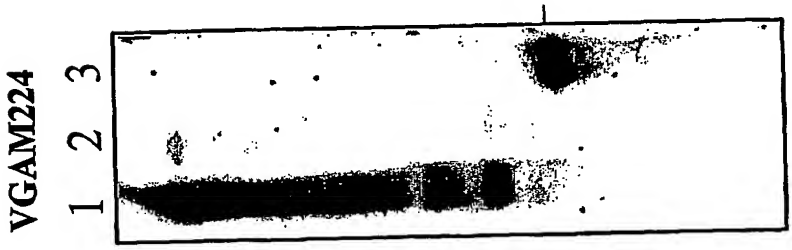


FIG.17B

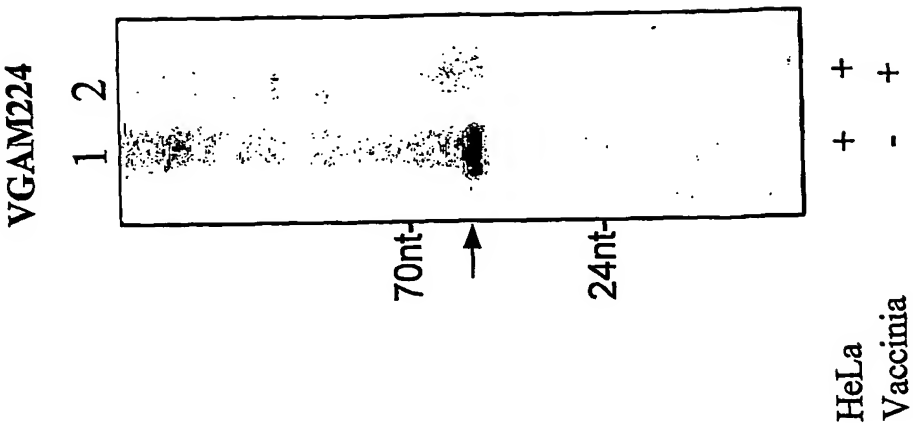



FIG.17C